JIDC | THE JOURNAL OF INFECTION IN DEVELOPING COUNTRIES

Original Article

Evaluation of the impact of Shigella virulence genes on the basis of clinical features observed in patients with shigellosis

Visnu Pritom Chowdhury[1,2,3], Ishrat Jahan Azmi[2], Md Ahshanul Haque[4], Mohammad Rafiqul Islam[1], Mahmuda Akter[2], Shahin Mahmud[2,5], Abu Syed Golam Faruque[4], Kaisar Ali Talukder[2,5]

[1] Department of Mathematics and Natural Sciences, BRAC University, Dhaka, Bangladesh
[2] Laboratory Sciences and Services Division, International Centre for Diarrhoeal Disease Research (icddr,b), Dhaka, Bangladesh
[3] Department of Biological Sciences, University of Delaware, Newark, DE, United States
[4] Nutrition and Clinical Services Division, International Centre for Diarrhoeal Disease Research (icddr,b), Dhaka, Bangladesh
[5] Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh

Abstract

Introduction: *Shigella* continues to cause significant morbidity and mortality each year, mostly in under-five children living in developing countries. We investigated the association between *Shigella* virulence genes and shigellosis.

Methodology: We randomly selected 61 *S. flexneri* strains isolated from patients in Bangladesh between 2009 and 2013, and evaluated the presence of 140 MDa large-virulence-plasmid (p140), and 22 virulence genes including *ipaH, ial, toxin*, and T3SS-related genes.

Results: We found p140 in 79% (n = 48) and *ipaBCD* in 90% (n = 55) strains, while seven strains were missing the p140. The prevalence of *ial* was 89%, *ipgC* and *ipgE* was 85%, and the prevalence for the remaining genes was < 85%. During the multivariate analysis, we found that instead of *sen*, the *Shigella* enterotoxin gene *set* along with several other virulence genes such as *ipgA, icsB, ipgB1, spa15*, and *mxiC*, were significantly influencing multiple clinical features relevant to shigellosis, including bloody stool, mucoid stool, and rectal straining.

Conclusions: We believe our model will help to determine the actual disease burden by directly looking for the genetic material in clinically suggestive patients, especially when detecting the causative organisms by traditional means is difficult.

Key words: *Shigella*; Shigellosis; T3SS; Virulence.

J Infect Dev Ctries 2022; 16(12):1835-1844. doi:10.3855/jidc.16783

(Received 05 May 2022 – Accepted 04 October 2022)

Introduction

The global incidence of shigellosis has reduced significantly through improved sanitation and safe water supply. However, several multi-center studies, including MAL-ED (The Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development) and GEMS (Global Enteric Multicenter Study) indicate that it continues to be one of the significant causes of morbidity and mortality every year, especially among children under 5 years of age in sub-Saharan Africa and South Asia [1–3]. Kotloff and her team estimated that in 1999 the global annual burden of *Shigella* infection was 164.7 million cases with 1.1 million deaths, out of which 163.2 million were from developing countries and 69% were children under five years of age [4]. In 2016, a study concluded that *Shigella* was the second leading cause of diarrheal mortality among all ages accounting for 212,438 deaths, which included 63,713 deaths in under-five children [5].

There are four species of *Shigella*: *S. dysenteriae* (n = 15), *S. flexneri* (n = 23), *S. boydii* (n = 20), and *S. sonnei* (n = 1) [6], and they are further classified into different serotypes [7]. Although *S. dysenteriae* type 1 is responsible for the most severe shigellosis, no incidence has been reported since 2004. We have observed a drift in the *Shigella* epidemiology over the last two decades. In just ten years, the prevalence of *S. sonnei* tripled from 8% (in 2000) to 27% (in 2010) [7]. However, for the last 35 years, *S. flexneri* has remained the most predominant species in Bangladesh.

The cellular pathogenesis and clinical presentation of dysenteric illness are mediated by multiple *Shigella* virulence factors including, p140, enterotoxins, and Type 3 Secretion system (T3SS) [8]. The p140 is a 230 kbp long virulence plasmid containing 100 genes with

a 31 kbp conserved "entry region" composed of 34 genes clustered in four groups. Group I contains the effector genes encoding the invasion plasmid antigen (Ipa). These Ipa proteins manipulate the host cell processes in favor of the pathogen [9]. Group II consists of *membrane expression of ipa* (*mxi*) and *surface presentation of ipa* (*spa*) genes, encoding Mxi and Spa proteins that are required for secretion of the Ipa and other effector proteins [10]. Group III has two regulatory genes, *virB* and *mxiE* that regulate the *T3SS* genes [11]. Group IV contains chaperone protein-coding genes that are required to stabilize the *T3SS* substrates within the bacterial cytoplasm [12]. However, all operations of *p140* are strictly controlled by the global regulatory elements described elsewhere [8].

The toxins are essential virulence factors needed for the survival of *Shigella* spp. Unlike the *ShET2* - encoded by the *sen* gene, located in p140, *Shiga toxin* (*Stx*) and *Shigella enterotoxin 1* (*ShET1*) are encoded in the chromosome. *Stx* is found in *S. dysenteriae* type 1 and enterohemorrhagic *E. coli* [13], while *ShET1* is found in *S. flexneri* [14]. However, *Stx* is responsible for the more severe complications, including hemolytic-uremic syndrome, and even death [15].

The T3SS - a needle-like virulence apparatus – has a base, needle, inner membrane export apparatus, cytosolic components, tip complex, and translocons. The base spans over both the outer membranes and is composed of two concentric rings. The outer ring is made of MxiD protein, and the inner ring is made of MxiG and MxiJ proteins [16]. The outer rod primarily consists of ~100 copies of MxiH protein arranged in a helical pattern. The inner rod is made of MxiI and this determines the needle length. The inner membrane export apparatus consists of five conserved proteins - MxiA, Spa24, Spa9, Spa29, and Spa40 - acting as a protein channel and facilitating the target proteins through the inner membrane [16]. The cytosolic components recruit, unfold, and transport the substrates after contact with the host cell, while a linker protein - MxiN - recruits the cytosolic components - e.g., ATPase - to the sorting platform [17]. The activation of the needle complex occurs following contact with the host cell, mediated by the tip complex - i.e., IpaB, IpaC, and IpaD [18] - passaging the effectors into the host cell and initiating pathogenesis. These cellular events can lead to serious illness in humans, ranging from diarrhea with tenesmus and high fever to severe life-threatening complications, including *Shigella* encephalopathy, hemolytic-uremic syndrome, and even death.

Shigella is easily transmittable through the fecal-oral route. With no effective vaccine [3], an infectious dose as low as 10-100 organisms [19], and with the emergence of multi-drug resistant *Shigella* spp. [20], it is wise to consider *Shigella* as a formidable public health threat that can wreak havoc on the health system in any densely populated developing tropical country, like Bangladesh. Since *Shigella* can remain undetected during the culture and sensitivity assay [21], alternative means should be sought to determine the actual disease burden and for evidence-based treatment for this pathogen. Therefore, we evaluated which *Shigella* virulence genes influence a particular clinical feature – individually or as a cluster – during shigellosis in humans.

Methodology

Ethical statement

This study was done on children under the age of five years and was part of a surveillance system aiming to understand the etiology and burden of diarrheal disease. The study was approved by the Research Review Committee and the Ethical Review Committee of icddr,b in December 2009. Caregivers of the children provided informed written consent before enrollment and collection of stool samples.

Bacterial isolates

Sixty-one clinical isolates of *S. flexneri* of different serotypes were obtained from the patients enrolled in the "Disease burden and etiologic agents of diarrhea patients visiting Kumudini Hospital, Mirzapur" study from 2009 to 2013. Different serotypes were present among these 61 *S. flexneri* strains: 1b (n = 5), 1c (n = 9), 2a (n = 16), 2b (n = 15), 3a (n = 8), 4 (n = 3), and 6 (n = 5). All these strains were isolated and characterized at the Enteric and Food Microbiology Laboratory of icddr,b following standard microbiological and biochemical methods [22]. A single colony of confirmed *Shigella* strain was grown in Tryptic Soy Broth (TSB) with 0.3% yeast extract and stored at -70 °C after adding 15% glycerol for further use. The strains that were used as plasmid standards (i.e., *E. coli* PDK-9, V-517, Sa, and R1) were collected from the Enteric and Food Microbiology Laboratory of icddr,b. The corresponding clinical features data was collected from the database of the aforementioned study at the beginning of this paragraph.

Serological Typing

We confirmed each strain using a commercially available antisera kit (Denka Seiken, Tokyo, Japan) and

monoclonal antibody (Reagensia AB, Stockholm, Sweden), specific for all *S. flexneri* type and group-factor antigens. We also sub-cultured the strains on MacConkey agar plates (Difco, Becton Dickinson & Company, Sparks, MD, USA). We performed the glass slide agglutination test after overnight incubation to confirm the serotype as described elsewhere [23].

Plasmid analysis

Plasmid DNA was processed according to the alkaline lysis method of Kado and Liu [24] with few modifications [25]. An isolated colony for each strain was inoculated in 1.5 mL Tryptic Soy broth with 0.3% yeast extract (TSBY) and incubated overnight at 37 °C on a water bath shaker. Later the cells were harvested by centrifugation and were suspended in 100 µL of solution I (40 mM Tris-Na acetate, 2 mM EDTA, pH 7.4). Then 200 µL of solution II (3% sodium dodecyl sulfate, 50 mM Tris, pH 12.9) was added and incubated at 55 °C for 1 hour. After incubation, an equal volume of solution III - phenol-chloroform-isoamyl alcohol [25:24:1] - was added and mixed carefully before

Table 1. List of primers used to identify *Shigella* virulence genes.

Primers	Oligonucleotide sequence (5' to 3')	Product size (bp)	Reference
ipaH F	TGGAAAAACTCAGTGCCTCT	423	[29]
ipaH R	CCAGTCCGTAAATTCATTCT		
ial F	CTGGATGGTATGGTGAGG	320	[30]
ial R	GGAGGCCAACAATTATTTCC		
set1A F	TCACGCTACCATCAAAGA	309	[27]
set1A R	TATCCCCCTTTGGTGGTA		
set1B F	GTGAACCTGCTGCCGATATC	147	[27]
set1B R	ATTTGTGGATAAAAATGACG		
sen F	ATGTGCCTGCTATTATTTAT	799	[27]
sen R	CATAATAATAAGCGGTCAGC		
virB F	GAAAAGTTGCGGTCTC	255	[31]
virB R	AGGTAATTCTCCGGCC		
ipaBCD F	GCTATAGCAGTGACATGG	500	[32]
ipaBCD R	ACGAGTTCGAAGCACTC		
ipgC F	CACCATGTCTTTAAATATCACCG	475	[33]
ipgC R	ATACTCCTTGATATCCTGAATTG		
ipgB1 F	CGGGATCCCATATAGGGGGTATCATG	650	[34]
ipgB1 R	GCGTCGACTTAATTTGTATTGCTTTG		
ipgA F	CGCGGATCCATGTGTCGCAAACTATATG	435	[35]
ipgA R	CCGGAATTCTTAGTTCACTTCTGAAGTG		
icsB F	CGCGGATCCATGAGCCTCAAAATTAGCAA	1568	[35]
icsB R	CCGGAATTCCTATATATTAGAATGAG		
ipgD F	TCGGCGTCAGAAGAGAAGTC	560	[36]
ipgD R	TATTAGCACATCATCATCAA		
ipgE F	CCACCGGTCGACTTAATACCCCTTCATTCTTCG	390	[37]
ipgE R	CCACGCGGATCCATGGAAGATTTAGCAGATGTT		
ipgF F	GCTTGAATTCGATTGTTGGGATAAGGCTGG	441	[38]
ipgF R	GAGCAAGCTTTTATTATATCCTTCGATTATTCTGCTTGCTC		
mxiH F	AGCGGATCCAGTGTTACAGTACCG	283	[39]
mxiH R	CGCGTCGACTGGATTATCTGAAGT		
mxiI F	CATGCCATGGTTTACATTTATCCAGTC	318	[39]
mxiI R	CCCAAGCTTAGACTTTAATAAAGTTTC		
mxiK F	GGCGTAATCAGAAGTGAG	575	[40]
mxiK R	TATACGCAGTGATTCAGC		
mxiE F	GCATAGCCATGCTGAAAAATG	249	This study
mxiE R	TCCGACACACCATAATGCTC		
mxiC F	GTAGGTGATGTATGCTTG	1067	[36]
mxiC R	GATCACTTTCATCTCCTG		
spa15 F	ACAGCCATTCAGCGATTACC	198	This study
spa15 R	AACGATGAAGCTCTACCAGCTC		
spa47 F	CTCGGGGCCAATCTTGGAAAC	237	[40]
spa47 R	CGGCCGACTTGCGTATGAAGA		
spa32 F	GCTCGCATGCCTTTTGGAGGATGAT	958	[39]
spa32 R	GGCCGGATCCAAGAACCATTTACT		
spa24 F	TTCTGATTCTGGCTTGATG	547	[41]
spa24 R	GTGTCTGCTCTTGGAGTTG		

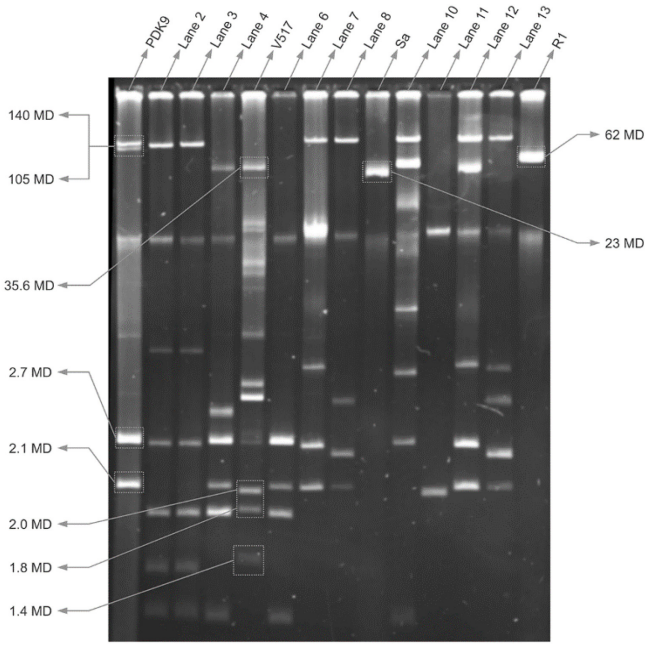
Gene name F (Forward primer), *Gene name* R (Reverse primer)

collecting the plasmid DNA by centrifugation. The plasmid DNA was separated by horizontal gel electrophoresis in a 0.7% agarose gel at room temperature at 100V (30 mA) in TBE (Tris-Borate-EDTA) buffer for approximately 3 hours. The slab gel was stained with ethidium bromide and the DNA band images were taken by a gel documentation system. The masses of unknown plasmid DNA were measured by comparing the mobilities with the known molecular weights [26]. The plasmids present in *E. coli* strains PDK-9 (140, 105, 2.7 and 2.1 MDa), V-517 (35.8, 4.8, 3.7, 3.4, 3.1, 2.0, 1.8 and 1.4 MDa), Sa (23 MDa) and R1 (62 MDa) were used as molecular weight standards.

PCR assay

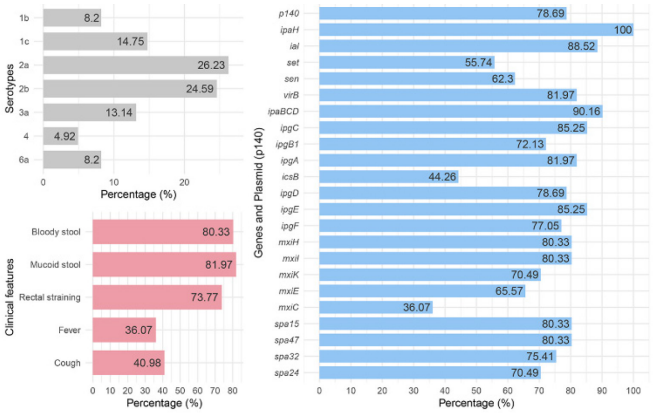
The detection of the following genes in *Shigella* isolates was carried out by polymerase chain reaction (PCR) assay: *ipaH, ial, set* (*Shigella enterotoxin 1*), *sen* (*Shigella enterotoxin 2*), and T3SS-related genes (*virB, ipaBCD, ipgC, ipgB1, ipgA, icsB, ipgD, ipgE, ipgF, mxiH, mxiI, mxiK, mxiE, mxiC, spa15, spa47, spa32 and spa24*) according to the procedure described

Figure 2. Plasmid DNA analysis of representative S. flexneri strains.



Gel electrophoresis of characterized plasmid DNA. Lane 1, *E. coli* PDK-9; Lane 2, K-4081 *S. flexneri* 6a; Lane 3, K-2336 *S. flexneri* 6a; Lane 4, K-3253 *S. flexneri* 1c; Lane 5, *E. coli* V-517; Lane 6, K-3127 *S. flexneri* 1b; Lane 7, K-2980 *S. flexneri* 2a; Lane 8, K-1502 *S. flexneri* 3a; Lane 9, *E. coli* Sa; Lane 10, K-632 *S. flexneri* 4; Lane 11, K-553 *S. flexneri* 4; Lane 12, K-102 *S. flexneri* 2b; Lane 13, K-6 *S. flexneri* 3b; Lane 14, *E. coli* R1. The position of the reference strains at 140 MDa, 105 MDa, 62 MDa, 35.6 MDa, 23 MDa, 2.7 MDa, 2.1 MDa, 2.0 MDa, 1.8 MDa and 1.4 MDa plasmid DNA shown with indicators.

Figure 1. The status of different serotypes, associated clinical features, the p140, and the virulence genes in *S. flexneri* isolates in this study.



previously (Table 1) [27]. However, PCR protocols for *spa15* and *mxiE* were developed in this study after designing the respective PCR primers, following instructions described elsewhere [28]. During PCR, *S. flexneri* 2a strain YSH6000 was used as the positive control, and *E. coli* ATCC 25922 lacking the p140 was used as the negative control.

Statistical analysis

We checked the difference in proportions of the explanatory variables (*Shigella* virulence genes) relative to the outcome variables (clinical features) and used the Chi-square tests of independence to check whether these differences are statistically significant. A p-value < 0.05 was considered statistically significant. When the expected counts were < 5, we used Fisher's exact test to determine significance. Next, we used binary logistic regression analysis on all response variables individually to assess whether the outcome variables were significantly ($p < 0.05$) influenced by individual or multiple explanatory variables. To avoid the multicollinearity effect, we merged several predictor variables into one, especially when one predictor variable – a gene – had a well-known dependent mechanism of action with another predictor variable – i.e., another gene. We used R (version 4.0.4) for all data analysis, and prepared the graphs and figures using Image Lab Version 5.2.1, Image J 1.48k, GNU Image Manipulation Program Version 2.8.10, and draw.io Version 11.3.0 software.

Results

Serological status

All the 61 isolates were *S. flexneri* consisting of different serotypes 1b (n = 5), 1c (n = 9), 2a (n = 16),

2b (n = 15), 3a (n = 8), 4 (n = 3), and 6a (n = 5) (Figure 1).

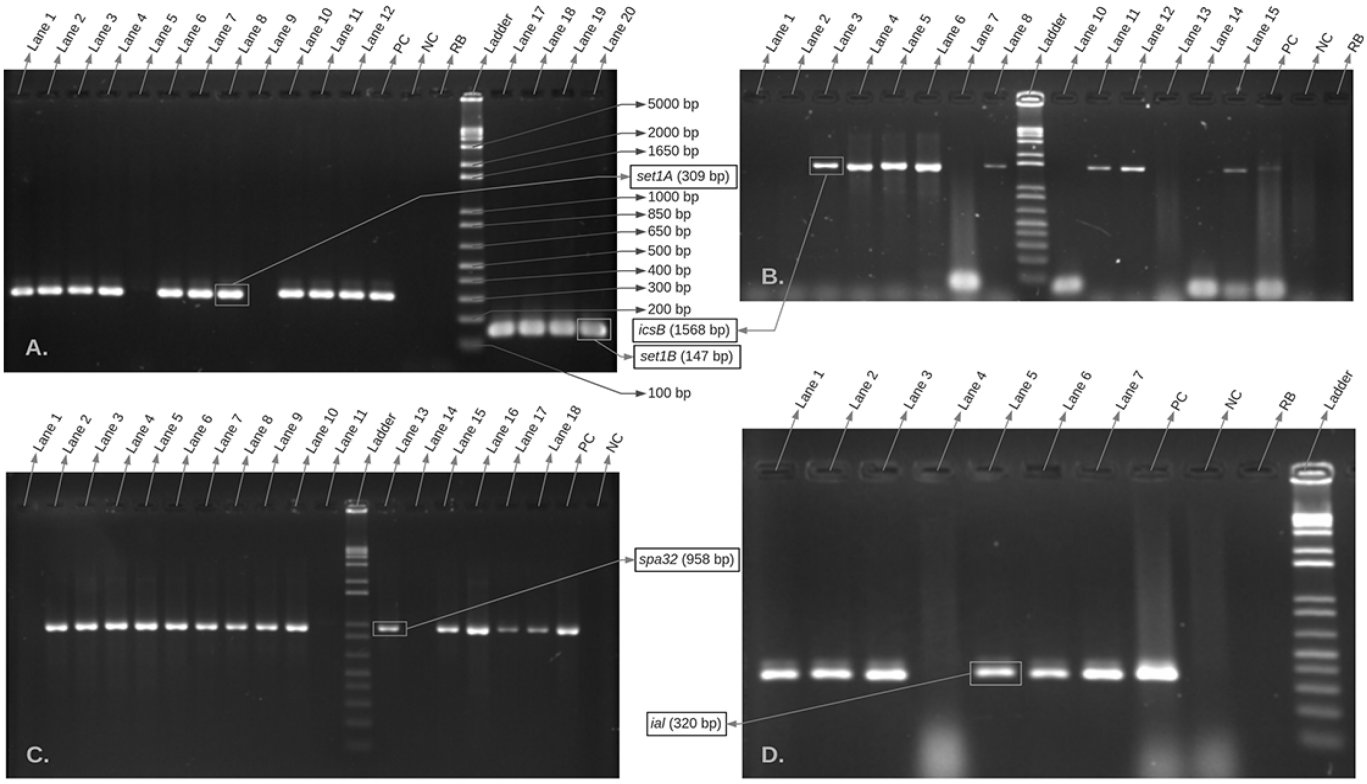
Status of plasmid profile

Out of the 61 isolates, 79% (48/61) were p140 positive. *S. flexneri* 2a (n = 15), 2b (n = 14), 3a (n = 8), Type-4 (n = 3), and 6a (n = 5) serotypes were present among these 48 strains. None of the 1b strains were p140 positive, while three 1c strains (3/9) were positive for p140. Small plasmids, approximately 2.7 and 2.1 MDa, were observed in around 94% strains, thus as core plasmids. Nonetheless, the overall plasmid population was heterogeneous (Figure 2).

Assessment of virulence genes

We analyzed the 61 *S. flexneri* strains to evaluate the presence of *ipaH, ial,* toxin (*set, sen*), and 18 different T3SS-related genes - *virB, ipaBCD, ipgC, ipgB1, ipgA, icsB, ipgD, ipgE, ipgF, mxiH, mxiI, mxiK, mxiE, mxiC, spa15, spa47, spa32* and *spa24* genes by PCR assay (Figures 1 and 3). All were found to be *ipaH* positive, while *ial* was found in 54 isolates. All of the 7 *ial* negative strains were also p140 negative, but 6 *ial* positive strains were found p140 negative. The *sen* gene was found in 38 strains and was absent in 23. In addition, 34 *S. flexneri* strains were *set* positive, and 27 were *set* negative. Out of the 27 *set* negative strains, 22

Figure 3. Gel electrophoresis of PCR products of representative *S. flexneri* strains.



Agarose gel electrophoresis of (**A**) *set1A*, *set1B* genes: Lane 1 – K1063 – *S. flexneri* 2a, Lane 2 – K1057 – *S. flexneri* 2a, Lane 3 – K1053 – *S. flexneri* 2b, Lane 4 – K1044 – *S. flexneri* 2b, Lane 5 – K842 – *S. flexneri* 1c, Lane 6 – K662 – *S. flexneri* 2b, Lane 7 – K658 – *S. flexneri* 2b, Lane 8 – K645 – *S. flexneri* 2b, Lane 9 – K632 – *S. flexneri* 4, Lane 10 – K629 – *S. flexneri* 2b, Lane 11 – K583 – *S. flexneri* 2b, Lane 12 – K570 – *S. flexneri* 2b, Lane 13 – Positive control (PC), Lane 14 – Negative control (NC), Lane 15 – Reagent blank (RB), Lane 16 – 1 kbp plus DNA ladder, Lane 17 – K1063 – *S. flexneri* 2a, Lane 18 – K1057 – *S. flexneri* 2a, Lane 19 – K1053 – *S. flexneri* 2b, Lane 20 – K1044 – *S. flexneri* 2b; (**B**) *icsB* gene: Lane 1 – K1044 – *S. flexneri* 2b, Lane 2 – K842 – *S. flexneri* 1c, Lane 3 – K662 – *S. flexneri* 2b, Lane 4 – K658 – *S. flexneri* 2b, Lane 5 – K649 – *S. flexneri* 2a, Lane 6 – K645 – *S. flexneri* 2b, Lane 7 – K632 – *S. flexneri* 4, Lane 8 – K629 – *S. flexneri* 2b, Lane 9 – 1 kbp plus DNA ladder, Lane 10 – K583 – *S. flexneri* 2b, Lane 11 – K570 – *S. flexneri* 2b, Lane 12 – K569 – *S. flexneri* 2b, Lane 13 – K553 – *S. flexneri* 4, Lane 14 – K425 – *S. flexneri* 2b, Lane 15 – K270 – *S. flexneri* 3a, Lane 16 – Positive control, Lane 17 – Negative control, Lane 18 – Reagent blank; (**C**) *spa32* gene: Lane 1 – K842 – *S. flexneri* 1c, Lane 2 – K662 – *S. flexneri* 2b, Lane 3 – K658 – *S. flexneri* 2b, Lane 4 – K649 – *S. flexneri* 2a, Lane 5 – K645 – *S. flexneri* 2b, Lane 6 – K632 – *S. flexneri* 4, Lane 7 – K629 – *S. flexneri* 2b, Lane 8 – K583 – *S. flexneri* 2b, Lane 9 – K570 – *S. flexneri* 2b, Lane 10 – K569 – *S. flexneri* 2b, Lane 11 – K553 – *S. flexneri* 4, Lane 12 – 1 kbp plus ladder, Lane 13 – K425 – *S. flexneri* 2b, Lane 14 – K270 – *S. flexneri* 3a, Lane 15 – K151 – *S. flexneri* 2b, Lane 16 – K102 – *S. flexneri* 2b, Lane 17 – K53 – *S. flexneri* 1c, Lane 18 – K6 – *S. flexneri* 3b, Lane 19 – Positive control, Lane 20 – Negative control; (**D**) *ial* gene: Lane 1 – K629 – *S. flexneri* 2b, Lane 2 – K583 – *S. flexneri* 2b, Lane 3 – K570 – *S. flexneri* 2b, Lane 4 – K553 – *S. flexneri* 4, Lane 5 – K270 – *S. flexneri* 3a, Lane 6 – K151 – *S. flexneri* 2b, Lane 7 – K53 – *S. flexneri* 1c, Lane 8 – Positive control, Lane 9 – Negative control, Lane 10 – Reagent blank, Lane 11 – 1 kbp plus ladder. The expected positions of the PCR product of genes are shown with the indicators. In each case, 1 kbp plus DNA ladder was used. YSH6000 *S. flexneri* 2a was used as positive control and *E. coli* ATCC 25922 was used as negative control.

were both *set1A* and *set1B* negative, and the remaining 5 were negative for either *set1A* or *set1B*. However, all *S. flexneri* 1b and 1c strains were *set* negative. On the other hand, *ipaBCD* was found in 90% of strains. The prevalence of *ipgC* and *ipgE* was 85% (52/61), *virB* and *ipgA* was 82% (50/61), *mxiK* and *spa24* was 70% (43/61), and *mxiH, mxiI, spa15, spa47* prevalence was 80% (49/61). The prevalence of the remaining genes (*ipgD* (n = 79), *ipgF* (n = 77), *spa32* (n = 75), *ipgB1* (n = 72), *mxiE* (n = 66), *icsB* (n = 44), and *mxiC* (n = 36) were variable. Although 13 strains were found to be *p140* negative, none of those strains were completely devoid of the T3SS genes and at least one T3SS-related gene was present in each of them. This count was much higher in two strains - K-1080 (*S. flexneri* 2a) and K-842 (*S. flexneri* 1c) (Figure 4).

Status of clinical information

Apart from the increased incidence of severe diarrheal disease (90.2%, n = 55), the clinical features observed in corresponding patients were: abdominal pain (85.2%, 52/61), bloody stool (80.3%, 49/61), mucoid stool (82.0%, 50/61), recorded temperature, i.e., fever (36%, 22/61), rectal straining (73.8%, 45/61), cough (41%, 25/61), vomiting (45.9%, 28/61), dehydration (21.3%, 13/61), and history of convulsion (3.2%, 2/61). The status of the corresponding clinical features observed in the p140 negative strains is shown in Figure 1.

Difference in proportion of virulence genes and features of shigellosis

We found several statistically significant ($p < 0.05$) differences in proportions among the *Shigella* virulence genes and relevant clinical features associated with shigellosis (Table 2). The presence of the *ial* gene was significantly associated with abdominal pain (p = 0.030). One of the enterotoxin genes – *set* – was strongly associated with clinical features including, mucoid stool, bloody stool, and rectal-straining (all $p <$ 0.01).

Factors associated with clinical features of shigellosis

Binary logistic regression analysis (Table 3, Figure 5), indicated that the presence of the *Shigella* enterotoxin gene – *set* – was associated with rectal-straining (OR: 25.150, 95% CI: 3.320, 592.211; p = 0.009) even after adjusting other covariates (listed below the table). The *Shigella* enterotoxin gene – *set* – also had a statistically significant association (OR: 6.00, 95% CI: 1.763, 24.487; p = 0.006) with rectal straining during the unadjusted regression analysis. Similarly, the toxin gene – *set* – was found to be associated with bloody stool (OR: 24.791, 95% CI: 3.989, 278.175; p = 0.002) and mucoid stool (OR: 11.190, 95% CI: 2.297, 88.727; p = 0.007), which was also found to be significant during the unadjusted analysis. However,

Figure 4. Status of virulence genes and clinical features associated with the p140 negative *S. flexneri* strains

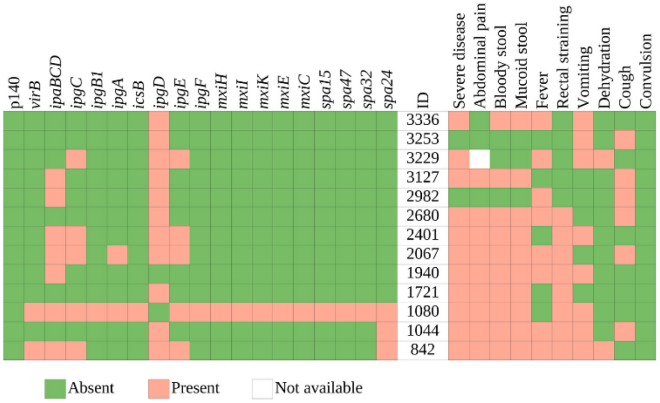


Figure 5. Comparison of regression models

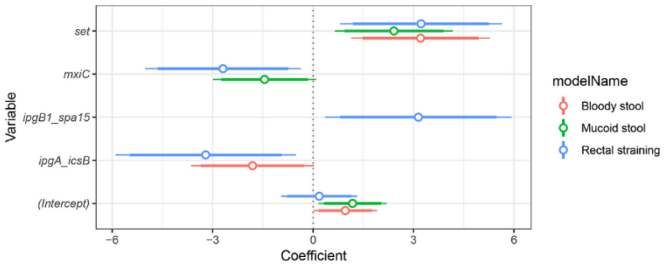


Table 2. Summary of descriptive statistics.

Independent variables (Genes)	Dependent variables (Clinical features)	Case (with clinical features) n (%)	Control (without clinical features) n (%)	*p*
ial	Abdominal pain	48 (92.31)	4 (57.14)	0.030*
	Mucoid stool	32 (64.00)	2 (18.18)	0.008*
set	Bloody stool	32 (65.31)	2 (16.67)	0.003
	Rectal straining	30 (66.67)	4 (25.00)	0.008
	Mucoid stool	32 (64.00)	2 (18.18)	0.008*
sen	Bloody stool	32 (65.31)	2 (16.67)	0.003
	Rectal straining	30 (66.67)	4 (25.00)	0.008

*Calculated by Fisher's exact test instead of Pearson's Chi square test.

the presence of the *icsB* and its chaperone *ipgA* was found to be significantly associated with bloody stool (OR: 0.164, 95% CI: 0.020, 0.897; p = 0.053) and rectal-straining (OR: 0.041, 95% CI: 0.001, 0.390; p = 0.020) only during the adjusted regression analysis. Both the *mxiC* (OR: 0.068, 95% CI: 0.003, 0.485; p = 0.023) and the *ipgB1* and its chaperone *spa15* (OR: 23.250, 95% CI: 2.112, 725.093; p = 0.027) were found to be associated with rectal-straining but the direction of association was opposite. Finally, although the enterotoxin gene – *set* – was found significantly associated with both fever (OR: 0.521, 95% CI: 0.177, 1.493; p = 0.227) and cough (OR: 0.333, 95% CI: 0.112, 0.944; p = 0.042) this association was only in the unadjusted models.

Discussion

We randomly selected 61 *S. flexneri* strains for plasmid profiling to detect the presence of the 140 MDa large virulence plasmid (p140) and performed PCR assay to determine the status of *Shigella* virulence genes, including *ipaH*, *ial*, toxin (*set, sen*), and T3SS-related genes (*virB, ipaBCD, ipgC, ipgB1, ipgA, icsB, ipgD, ipgE, ipgF, mxiH*, *mxiI, mxiK, mxiE, mxiC, spa15, spa47, spa32, spa24*). We also designed the primers for *mxiE* (a T3SS regulator) and *spa15* (a T3SS chaperone) genes with primer3plus and confirmed their validity when the predicted product size correlated with the band size of the PCR product during gel electrophoresis analysis.

Even though the T3SS related genes are located inside the 31 kbp "entry region" of p140 [8], none of these p140-negative strains were entirely devoid of the T3SS-related genes (Figure 4). We found 6 of such T3SS-related genes in K-842 (*S. flexneri* 1c) and 17 in K-1080 (*S. flexneri* 2a). We observed multiple clinical features in the patients from whom these p140-negative strains were isolated (Figure 4). A higher proportion of the clinical traits analyzed in this study were found in K-1080 (*S. flexneri* 2a) and K-842 (*S. flexneri* 1c), including abdominal pain, bloody mucoid stool, fever, rectal straining, vomiting, and an increased severity of diarrheal disease. These strains were pathogenic even though they were missing p140. Therefore, we suspect that these T3SS-related genes might be integrated elsewhere within the bacterial genome, probably mediated by transposons or insertion sequences [42].

During the bivariate analysis, we found that the *Shigella* enterotoxin gene *set* was strongly associated with mucoid stool, bloody stool, and rectal straining (p values < 0.01). However, the chromosomal *set* gene is known to have two domains - *set1A* and *set1B* - encoding ShET1A and ShET1B proteins. The ShET1A is responsible for the secretory activity, while ShET1B causes irreversible binding of the toxin to the enterocyte receptor [43]. In contrast, the plasmid-borne *sen* gene encoding ShET2 (*Shigella* enterotoxin 2) causes epithelial inflammation by contributing to the release of Interleukin-8 (IL-8) from the gut epithelium [44]. IL-8 is a chemokine that attracts and activates the neutrophil at the inflammatory region, causing injury to the nearby cells, including the rectal epithelium [45]. We also found *ial* to be significantly associated with abdominal pain (p = 0.030), while this gene is known to direct epithelial cell penetration by the pathogen [46].

Since these 22 T3SS-related genes may be present on the same plasmid, we wanted to identify which genes (individually or in combination with other genes) are mostly associated with a particular clinical outcome of shigellosis. To do that, we used binary logistic regression analysis after merging several variables into one depending on their mechanistic relationship to avoid the multicollinearity effect (Table 3, Figure 5). In the multivariate analysis, we found that the presence of rectal straining was significantly influenced by the simultaneous presence of the enterotoxin *set* gene, the *icsB* which camouflages *IcsA* from the autophagic host defense system, its chaperone IpgA which is required

Table 3. Summary of binary logistic regression analysis.

Outcome variables	Predictor variables	Unadjusted		Adjusted	
		OR (95% CI)	*p*	OR (95% CI)	*p*
Fever	*set* + *sen*	0.521 (0.177, 1.493)	0.227	-	-
Bloody stool	*set* + *sen*	9.412 (2.175, 65.988)	0.007	24.791 (3.989, 278.175)	0.002
	ipgA + *icsB*	0.750 (0.207, 2.715)	0.656	0.164 (0.020, 0.897)	0.053
Cough	*set* + *sen*	0.333 (0.112, 0.944)	0.042	-	-
Mucoid stool	*set* + *sen*	8 (1.820, 56.386)	0.013	11.190 (2.297, 88.727)	0.007
	mxiC	0.392 (0.099, 1.485)	0.167	0.235 (0.045, 1.048)	0.065
Rectal straining	*set* + *sen*	6.00 (1.763, 24.487)	0.006	25.150 (3.320, 592.211)	0.009
	ipgA + *icsB*	0.731 (0.229, 2.321)	0.591	0.041 (0.001, 0.390)	0.020
	mxiC	0.452 (0.138, 1.457)	0.181	0.068 (0.003, 0.485)	0.023
	ipgB1 + *spa15*	2.722 (0.799, 9.290)	0.105	23.250 (2.112, 725.093)	0.027

List of adjusted variables: *ial*, *set* + *sen, virB*, *ipaBCD* + *ipgC*, *ipgB1* + *spa15*, *ipgA* + *icsB*, *ipgD* + *ipgE*, *ipgF*, *mxiH* + *mxiI, mxiK, mxiE, mxiC, spa47*, *spa32* + *spa24*.

for the stability and secretion of *IcsB* [35], the *mxiC* which is a *Shigella* T3SS substrate that prevents premature effector secretion [47], the *ipgB1* which induces membrane ruffling to increase invasiveness, and its chaperone *spa15*, that stabilize the *IpgB1* [34,48]. We also found that bloody stool was significantly associated with the presence of the toxin *set* gene, and the *icsB* and its chaperone *ipgA*, while the mucoid-stool was found to be associated with the enterotoxin *set* gene and another T3SS substrate gene *mxiC* ($p < 0.10$). Despite the *set* gene's known role in secretory activity [14,43], in this study, we found the *set* gene alone to be associated with invasive clinical features including bloody stool, mucoid stool, and rectal straining, which was previously not known and hence requires further investigation (Tables 2 and 3, Figure 5). Several studies have reported that shigellosis is associated with convulsion [49]. We found no significant association, probably because we had only two convulsion cases.

We have several limitations in our study. Our sample size was small, we could not sequence the *Shigella* strains, and we did not analyze the bacterial strains immediately after isolating pathogens from the patients. Therefore, the possibility remains that the plasmid and the genes might be lost during storage. There is also the possibility of co-infections in patients, and we do not have adequate information to analyze this aspect further. In addition, we did not have *Shigella* and non-*Shigella* cases to compare the presence of these genes in different pathological conditions.

Conclusions

In conclusion, we found that the *Shigella* enterotoxin gene *set* and several other virulence genes such as *ipgA, icsB*, *ipgB1*, *spa15,* and *mxiC*, were significantly associated with multiple shigellosis-relevant clinical features, including bloody stool, mucoid stool, and rectal straining in separate models. Although our model is based on a small dataset with considerable limitations, we believe that such a model could help us determine the actual disease burden by directly looking for the genetic material (i.e., *ipaH*, even though this gene is also found in enteroinvasive *Escherichia coli*), especially when a patient is clinically suggestive but still culture negative.

Acknowledgements

We are grateful to the core donors of icddr,b – the Government of the People's Republic of Bangladesh, Global Affairs Canada (GAC), Swedish International Development Cooperation Agency (SIDA), and the Department for International Development (UK AID) – for providing unrestricted support to icddr,b for its operations and research.

References

1. Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM (2018) Shigellosis. Lancet 391: 801–812.
2. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acácio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382: 209–222.
3. Mani S, Wierzba T, Walker RI (2016) Status of vaccine research and development for *Shigella*. Vaccine 34: 2887–2894.
4. Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM (1999) Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ 77: 651–666.
5. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houpt ER, Kang G, Kotloff KL, Levine MM, Luby SP, MacLennan CA, Pan WK, Pavlinac PB, Platts-Mills JA, Qadri F, Riddle MS, Ryan ET, Shoultz DA, Steele AD, Walson JL, Sanders JW, Mokdad AH, Murray CJL, Hay SI, Reiner RC (2018) Morbidity and mortality due to *Shigella* and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990–2016. Lancet Infect Dis 18: 1229–1240.
6. Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, Antonio M, Hossain A, Mandomando I, Ochieng JB, Oundo JO, Qureshi S, Ramamurthy T, Tamboura B, Adegbola RA, Hossain MJ, Saha D, Sen S, Faruque ASG, Alonso PL, Breiman RF, Zaidi AKM, Sur D, Sow SO, Berkeley LY, O'Reilly CE, Mintz ED, Biswas K, Cohen D, Farag TH, Nasrin D, Wu Y, Blackwelder WC, Kotloff KL, Nataro JP, Levine MM (2014) *Shigella* isolates from the global enteric multicenter study inform vaccine development. Clin Infect Dis 59: 933–941.
7. Talukder KA, Azmi IJ (2012) Population genetics and molecular epidemiology of *Shigella* species. In Shah M. Faruque, editor. Foodborne and waterborne bacterial pathogens: epidemiology, evolution and molecular biology. UK: Caister Academic Press. 63–76.
8. Schroeder GN, Hilbi H (2008) Molecular pathogenesis of *Shigella* spp: controlling host cell signaling, invasion, and death by type III secretion. Clin Microbiol Rev 21: 134–156.
9. Buysse JM, Stover CK, Oaks EV, Venkatesan M, Kopecko DJ (1987) Molecular cloning of invasion plasmid antigen (*ipa*)

genes from *Shigella flexneri*: analysis of *ipa* gene products and genetic mapping. J Bacteriol 169: 2561–2569.
10. Venkatesan MM, Buysse JM, Oaks EV (1992) Surface presentation of *Shigella flexneri* invasion plasmid antigens requires the products of the *spa* locus. J Bacteriol 174: 1990–2001.
11. Dorman CJ, Porter ME (1998) The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. Mol Microbiol 29: 677–684.
12. Parsot C, Hamiaux C, Page A-L (2003) The various and varying roles of specific chaperones in type III secretion systems. Curr Opin Microbiol 6: 7–14.
13. Bergan J, Dyve Lingelem AB, Simm R, Skotland T, Sandvig K (2012) Shiga toxins. Toxicon 60: 1085–1107.
14. Faherty C, Harper JM, Shea-Donohue T, Barry EM, Kaper JB, Fasano A, Nataro JP (2012) Chromosomal and plasmid-encoded factors of *Shigella flexneri* induce secretogenic activity ex vivo. PLoS One 7: e49980.
15. Kitov PI, Sadowska JM, Mulvey G, Armstrong GD, Ling H, Pannu NS, Read RJ, Bundle DR (2000) Shiga-like toxins are neutralized by tailored multivalent carbohydrate ligands. Nature 403: 669–672.
16. Galán JE, Lara-Tejero M, Marlovits TC, Wagner S (2014) Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. Annu Rev Microbiol 68: 415–438.
17. Akeda Y, Galán JE (2005) Chaperone release and unfolding of substrates in type III secretion. Nature 437: 911–915.
18. Blocker AJ, Deane JE, Veenendaal AKJ, Roversi P, Hodgkinson JL, Johnson S, Lea SM (2008) What's the point of the type III secretion system needle? Proc Natl Acad Sci U S A 105: 6507–6513.
19. DuPont HL, Levine MM, Hornick RB, Formal SB (1989) Inoculum size in shigellosis and implications for expected mode of transmission. J Infect Dis 159:1126–1128.
20. Azmi IJ, Khajanchi BK, Akter F, Hasan TN, Shahnaij M, Akter M, Banik A, Sultana H, Hossain MA, Ahmed MK, Faruque SM, Talukder KA (2014) Fluoroquinolone resistance mechanisms of *Shigella flexneri* isolated in Bangladesh. PLoS One 9: e102533.
21. Pavlinac PB, Platts-Mills JA, Tickell KD, Liu J, Juma J, Kabir F, Nkeze J, Okoi C, Operario DJ, Uddin J, Ahmed S, Alonso PL, Antonio M, Becker SM, Breiman RF, Faruque ASG, Fields B, Gratz J, Haque R, Hossain A, Hossain MJ, Jarju S, Qamar F, Iqbal NT, Kwambana B, Mandomando I, McMurry TL, Ochieng C, Ochieng JB, Ochieng M, Onyango C, Panchalingam S, Kalam A, Aziz F, Qureshi S, Ramamurthy T, Roberts JH, Saha D, Sow SO, Stroup SE, Sur D, Tamboura B, Taniuchi M, Tennant SM, Roose A, Toema D, Wu Y, Zaidi A, Nataro JP, Levine MM, Houpt ER, Kotloff KL (2020) The clinical presentation of culture-positive and culture-negative, quantitative polymerase chain reaction (qPCR)-attributable shigellosis in the global enteric multicenter study and derivation of a *Shigella* severity score: implications for pediatric *Shigella* vaccine trials. Clin Infect Dis 73: e569-e579.
22. World Health Organization (2003) Manual for the laboratory identification and antimicrobial susceptibility testing of bacterial pathogens of public health importance in the developing world: *Haemophilus influenzae*, *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, *Salmonella* serotype Typhi, *Shigella*, and *Vibrio cholerae*. World Health Organization 359 p.
23. El-Gendy A, El-Ghorab N, Lane EM, Elyazeed RA, Carlin NI, Mitry MM, Kay BA, Savarino SJ, Peruski LF (1999) Identification of *Shigella flexneri* subserotype 1c in rural Egypt. J Clin Microbiol 37: 873–874.
24. Kado CI, Liu ST (1981) Rapid procedure for detection and isolation of large and small plasmids. J Bacteriol 145: 1365–1373.
25. Talukder KA, Islam MA, Dutta DK, Hassan F, Safa A, Nair GB, Sack DA (2002) Phenotypic and genotypic characterization of serologically atypical strains of *Shigella flexneri* type 4 isolated in Dhaka, Bangladesh. J Clin Microbiol 40: 2490–2497.
26. Haider K, Huq MI, Talukder KA, Ahmad QS (1989) Electropherotyping of plasmid DNA of different serotypes of *Shigella flexneri* isolated in Bangladesh. Epidemiol Infect 102: 421–428.
27. Vargas M, Gascon J, Anta MTJD, Vila J (1999) Prevalence of *Shigella* enterotoxins 1 and 2 among *Shigella* strains isolated from patients with traveler's diarrhea. J Clin Microbiol 37: 3608–3611.
28. Dieffenbach CW, Lowe TM, Dveksler GS (1993) General concepts for PCR primer design. PCR Methods Appl 3: S30-37.
29. Talukder KA, Islam MA, Khajanchi BK, Dutta DK, Islam Z, Safa A, Alam K, Hossain A, Nair GB, Sack DA (2003) Temporal shifts in the dominance of serotypes of *Shigella dysenteriae* from 1999 to 2002 in Dhaka, Bangladesh. J Clin Microbiol 41: 5053–5058.
30. Frankel G, Giron JA, Valmassoi J, Schoolnik GK (1989) Multi-gene amplification: simultaneous detection of three virulence genes in diarrhoeal stool. Mol Microbiol 3: 1729–1734.
31. Berlutti F, Casalino M, Zagaglia C, Fradiani PA, Visca P, Nicoletti M (1998) Expression of the virulence plasmid-carried *apyrase* gene (*apy*) of enteroinvasive *Escherichia coli* and *Shigella flexneri* is under the control of H-NS and the *VirF* and *VirB* regulatory cascade. Infect Immun 66: 4957–4964.
32. Faruque SM, Khan R, Kamruzzaman M, Yamasaki S, Ahmad QS, Azim T, Nair GB, Takeda Y, Sack DA (2002) Isolation of *Shigella dysenteriae* Type 1 and *S. flexneri* strains from surface waters in Bangladesh: comparative molecular analysis of environmental *Shigella* isolates versus clinical strains. Appl Environ Microbiol 68: 3908–3913.
33. Lane TW, Gantt RW (2004) Rapid detection of biothreat agents based on cellular machinery. Sandia National Laboratories (SNL), Albuquerque, NM, and Livermore, CA (United States). Available: https://www.osti.gov/biblio/920835. Accessed 2 December 2022.
34. Ohya K, Handa Y, Ogawa M, Suzuki M, Sasakawa C (2005) IpgB1 is a novel *Shigella* effector protein involved in bacterial invasion of host cells: its activity to promote membrane ruffling via *rac1* and *cdc42* activation. J Biol Chem 280: 24022–24034.
35. Ogawa M, Suzuki T, Tatsuno I, Abe H, Sasakawa C (2003) IcsB, secreted via the type III secretion system, is chaperoned by IpgA and required at the post-invasion stage of *Shigella* pathogenicity. Mol Microbiol 48: 913–931.
36. Lan R, Lumb B, Ryan D, Reeves PR (2001) Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. Infect Immun 69: 6303–6309.
37. Niebuhr K, Jouihri N, Allaoui A, Gounon P, Sansonetti PJ, Parsot C (2000) IpgD, a protein secreted by the type III

secretion machinery of *Shigella flexneri*, is chaperoned by IpgE and implicated in entry focus formation. Mol Microbiol 38: 8–19.
38. Zahrl D, Wagner M, Bischof K, Bayer M, Zavecz B, Beranek A, Ruckenstuhl C, Zarfel GE, Koraimann G (2005) Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. Microbiology 151: 3455–3467.
39. Magdalena J, Hachani A, Chamekh M, Jouihri N, Gounon P, Blocker A, Allaoui A (2002) *Spa32* regulates a switch in substrate specificity of the type III secreton of *Shigella flexneri* from needle components to Ipa proteins. J Bacteriol 184: 3433–3441.
40. Jouihri N, Sory M-P, Page A-L, Gounon P, Parsot C, Allaoui A (2003) MxiK and MxiN interact with the Spa47 ATPase and are required for transit of the needle components MxiH and MxiI, but not of Ipa proteins, through the type III secretion apparatus of *Shigella flexneri*. Mol Microbiol 49: 755–767.
41. Yang J, Wang J, Chen L, Yu J, Dong J, Yao Z-J, Shen Y, Jin Q, Chen R (2003) Identification and characterization of simple sequence repeats in the genomes of *Shigella* species. Gene 322: 85–92.
42. Pilla G, McVicker G, Tang CM (2017) Genetic plasticity of the *Shigella* virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences. PLoS Genet 13: e1007014.
43. Fasano A, Noriega FR, Liao FM, Wang W, Levine MM (1997) Effect of *Shigella enterotoxin 1* (*ShET1*) on rabbit intestine in vitro and in vivo. Gut 40: 505–511.
44. Farfán MJ, Toro CS, Barry EM, Nataro JP (2011) Shigella enterotoxin-2 is a type III effector that participates in *Shigella*-induced interleukin 8 secretion by epithelial cells. FEMS Immunol Med Microbiol 61: 332–339.
45. Bickel M (1993) The role of interleukin-8 in inflammation and mechanisms of regulation. J Periodontol 64: 456–460.
46. Ménard R, Prévost MC, Gounon P, Sansonetti P, Dehio C (1996) The secreted Ipa complex of *Shigella flexneri* promotes entry into mammalian cells. Proc Natl Acad Sci USA 93: 1254–1258.
47. Martinez-Argudo I, Blocker AJ (2010) The *Shigella* T3SS needle transmits a signal for MxiC release, which controls secretion of effectors. Mol Microbiol 78: 1365–1378.
48. Lillington JED, Lovett JE, Johnson S, Roversi P, Timmel CR, Lea SM (2011) *Shigella flexneri* Spa15 crystal structure verified in solution by double electron electron resonance. J Mol Biol 405: 427–435.
49. Khan WA, Dhar U, Salam MA, Griffiths JK, Rand W, Bennish ML (1999) Central nervous system manifestations of childhood shigellosis: prevalence, risk factors, and outcome. Pediatrics 103: E18.

Corresponding author

Professor Kaisar Ali Talukder, PhD
Department of Biotechnology and Genetic Engineering
Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh
Phone: +880 1715 105709
Email: katalukdar@yahoo.com

Conflict of interests: No conflict of interests is declared.